

## DOCUMENT RESUME

ED 170 779

CS 204 835

AUTHOR Miller, Susan  
TITLE Test Scores, Course Grades, and Bottom Lines.  
PUB DATE Apr 79  
NOTE 11p.; Paper presented at the Annual Meeting of the Conference on College Composition and Communication (30th, Minneapolis, Minnesota, April 5-7, 1979)

EDRS PRICE MF01/PC01 Plus Postage.  
DESCRIPTORS \*Composition (Literary); \*English Instruction; \*Evaluation Criteria; \*Grading; Higher Education; \*Student Evaluation; \*Writing Skills

IDENTIFIERS \*Reader Response

## ABSTRACT

There is often little correlation between objective tests of writing or writing components and grading by teachers. The technology that can be applied to student writing evaluation lags behind a reasoned rhetorical explanation of test results. Evaluations of writing are inadequate unless they are interpreted within a rhetorical context that includes who the writer is and for whom that writer is writing. Questions of quality thereby become part of a complicated context of interaction between specific writers and specific readers. The motivation and attitudes of both contributors to the meaning of a text ought not to be ignored; therefore, it is important to devise new descriptions of the affective as well as the cognitive aspects of writing and reading and to account for the predispositions of writer and reader. An interpretive theory informed by measurements of variables in dynamic, organic situations is needed. (TJ)

\*\*\*\*\*  
\* Reproductions supplied by EDRS are the best that can be made \*  
\* from the original document. \*  
\*\*\*\*\*

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

TO THE EDUCATIONAL RESOURCES  
INFORMATION CENTER (ERIC)

Susan Miller  
Department of English  
University of Wisconsin-  
Milwaukee

## TEST SCORES, COURSE GRADES, AND BOTTOM LINES

Because I implement the required writing proficiency examinations for three of the colleges of the University of Wisconsin-Milwaukee while coordinating eight elective writing courses, I also explain, justify, and occasionally defend programmatic, course, or individual evaluations of writing. And because I do so, I am increasingly disquieted by our discipline's coalescing approach to valid and reliable evaluation. The news that Johnny couldn't write was followed, appropriately, by a rash of testing programs, an NCTE resolution about the nature of supportable tests, and some hard questions among ourselves about what we mean when we say that students write either well or poorly. In the past year two publications, O'Dell and Cooper's Evaluating Writing, and the Spring, 1978, issue of Basic Writing, have even further led programs and individual teachers toward new methods of evaluation. We have now what someone reared on "write about the parking situation and he'll flunk you" methods of grading might take for a complete inventory of trustworthy, fair methodologies.

What I find, however, is that although we have new words, the music remains the same; the hard questions that are raised about writing evaluations linger on. We still have no theory of evaluation, no philosophy of interpretation, in which to embed the results of our tests. Whether working as researchers in large programs or as teachers explaining a grade on a paper to one student, we have no way to connect most of the newly valid and reliable results of evaluation to the rhetorical situation of an author and a reader. Without a theory, we unknowingly invite attack, and will, paradoxically by virtue of our new methods, continue to do so unless we question and diversify their results.

- 2 -

For instance, administrators of programs, people in the situation I faced when trying to convince Ohio State University to begin an expensive but essential remedial program for those 2,000 open admissions students who scored below 15 on the English ACT test each year, may attempt, as I did, to show that students need additional instruction by showing that they cannot benefit from available instruction. The method adopted in such situations, demonstrating that trained readers of anonymous samples agree that a large percentage of student writers cannot achieve an agreed-upon minimal standard, may (as it was at Ohio State) be successful. And then questions about prior and continuing methods of evaluation begin:

When a remedial program required for 1/4 of the students begins, why is there no rise in the grade distribution for the regular freshman course, even when course standards and objectives remain the same? Why were 1/4 of the freshman students required to take remedial courses if 40% of the class were already getting A's and B's in required freshman English? Why is there, in sum, no discernible relationship between a teacher's judgment about a course grade and valid and reliable holistic measurements of writing samples?

In a second context, we receive attacks that our demonstrations in the name of program development do not invite. For instance, what does a teacher reply to Rosemary Hake and Joseph William's Chicago demonstration that although English teachers trained in holistic reading teach in class that good writing has a verbal (active) style, they actually rate higher papers written in the less readable but more ponderous nominal style associated with bureaucrats and academic research?

Or, in yet a third context, what do we say about the recent research showing that essays written by recognized professional writers inserted into

- 3 -

stacks of student essays on the same topic were judged by (again) trained holistic raters to be poorer than the student themes? How do we make this result congruent with Kellogg Hunt's standard evaluation that clearly places the syntactic maturity of recognized professional writers two steps above 12th graders? If we "know" that professionals write a more syntactically mature prose than students, would we ever credit even trained teachers' ratings of and in this case preference for student papers? Without rejecting syntactic maturity as a developmental goal, how might we explain the teachers' preferences?

And in a fourth, final, immediate, and perhaps most important context, what do we tell a student who correctly identified the use of the exclamation point on an objective test, but whose teachers has just circled all of his definitely correct uses of that mark on an essay for class? No one told the student that the most important rule about the exclamation mark is, in practice, to use it sparingly; someone lowered his grade. Students who do learn handbook rules about punctuation, diction, or usage may think, justifiably, that our class evaluations of their writing are rarely rule-governed.

We need to reconsider our interpretations of such judgments, especially of valid and reliable judgments, if we are to confront these persistent questions. The worst is true. Even with, or perhaps especially with, valid and reliable methods of evaluation, we appear to contradict ourselves. We wave Strunk and White at students' essays and give high grades for prose they despised. We appear to grade writing in class on the basis of unreliable or invalid procedures. We only believe in exclamation points when it suits us to. We beg for money to instruct students we at one minute call remedial and in the next give A's and B's.

- 4 -

A coherent theory of interpretation stands up to these criticisms as reliance on methods or tools of evaluation cannot. Our technology--the norms, statistical verifications, bar graphs, matrices, primary features, and relative readability indices now available to apply to student writing--lags behind our philosophy, our reasoned rhetorical explanation of test results. In order to show how far these tools lag behind our needs, let me review the principles and premises that inform them:

1. Evaluations of writing, whether they judge or describe, are measurement.
2. Measurements are always and only relative.
3. Formal, accepted measurements of writing may indicate quality relative to two perceptible variables: time, and frequency. Time elements may be identified and measured: the time in which a piece was written, the time it took to read the piece, the grade-level or age of the writer, the amount of instructional time given before or after the writing sample or other test was assigned, the number of years experience of the teacher, the politics of the particular decade in which the same was evaluated--all of these temporal elements may apply in evaluations of writing.
4. The frequency of the grammatical, syntactic, stylistic, or organizational markers within a piece of discourse are the other quantifiable variables measured by standard evaluative methods. The per/discourse unit numbers of errors, of t-units, of propositions, of generalizations, of transitions, of abstractions, of various embeddings, of allusions, of metaphors--all may be counted.
5. Currently touted methods of evaluating writing may aim to predict or count these temporal or frequency-related characteristics of writing in order to define one of the following:

- 5 -

1. Changes in student writing (either by individuals or groups) relative to time;
2. Appropriate instructional placement relative to frequency counts in the writing of other students, and
3. Achievement relative to previously defined syntactic, course performance, minimal competency, or proficiency standards.

The key term in each of these admissible aims of measurement, is, of course, "relative." Our tools evaluate writing in relationship to only time and frequency; even holistic readings judge writing relative to the frequency-related impressions of readers of one population of essays, and the rate of reader agreement in such evaluations is carefully controlled by training readers about what to look for.

This sort of evaluative technology, however it may be validated, wars with the processes of reading and writing a text. I would be the last person to suggest that whimsical, subjective responses to student writing should once more reign with elitist arbitrariness; "I just don't like it" is not, I know, an alternative to placement or proficiency based on mean scores and controlled test situations. Nonetheless, the results of newly controlled evaluations of writing are, as now described and practiced, inadequate unless they are interpreted within a rhetorical context. We are embracing evaluative tools that by their nature must ignore the situation of the reader or of the writer, and in doing so we are newly opening ourselves to warped methods of instruction tied to such evaluations.

To demonstrate what I mean, let us return to the recent research about trained readers' preferences for student over professional writing and try to

- 6 -

construe possible interpretations of this investigation. Teachers were trained to read papers holistically in order to rate them against other papers in a sample population. When these teachers then rated student writers higher than professional writers writing about the same topic, a number of conclusions might be drawn. It might be possible, for example, to say that teachers can't judge the quality of prose, even after training. Or to say that teachers prefer youthful points of view on any topic. Or to say that any group of readers will respond most favorably to writing closest to the norm in a sample--that the training in fact ruined their ability to read by defining the terms of success as relative to most of the writing in a limited number of samples.

But a rhetorician, in fact Aristotle, supplied yet another theory that would account for the apparent inaccuracy of, and thus lack of trust of, these readers. Aristotle knew that we might often discover teachers grading what they thought to be student writing lower if it omitted exemplification and made points without illustrations to back them up. The Rhetoric (2.21), explicitly says that maxims--broad generalizations--are suitable evidence for speakers of mature years, but that young speakers should not use generalities where they lack experience. The readers in this study were accurate and consistent with a rhetoric-based reading theory, the essential element omitted from a holistic judgment based on relative time and frequency. These readers responded as all readers do, on the basis of expectations about the writers, not exclusively, nor even primarily, on the basis of expectations about a text.

Reading is always a situational rather than text-centered act. Readers, like writers, make meaning, read to some purpose. Their mood, the recompence for reading--in general, their motives--will determine their responses to any text just as a writer's intentions and sense of purpose will define the content, tone, and voice of writing. Just as writers have more or less experience with a



- 7 -

variety of aims and modes of discourse, and thus may be skilled in one form, format, or audience but not another, so also are readers more or less familiar with particular sorts of writing and more or less used to being the audience for certain intentions. They are prejudicially well or ill predisposed to certain authors or groups of authors. And their familiarity with the diction, syntax, and historical or invented examples of the author will determine their ability to identify with that author, to construct the meaning that was intended when composing.

This description of the reading process does not define "relative readability" as behaviorists would, as a matter of processing time depending on given contexts. It does not say, as Diederich did, that readers are variously predisposed toward certain features of texts. Nor does it mean to suggest that readers are consistent or capable of classification according to one or another habitual reading pattern. Instead it places questions of quality in a complicated context of interaction between specific writers and specific readers who have many writing and reading processes. And neither is this situational model of the reading process only a theoretical abstraction. It explains why evaluators want to know "who wrote this?" before they read as often as visitors ask the sex of a new baby before they speak to it. It explains why students who do poorly on writing samples may get high grades in courses, the grades of which measure the quality of relationship between reader and writer, not only the quality of an anonymous text.

Even the most enlightened and helpful evaluators of student writing, Lee O'Dell and Charles Cooper, suggest evaluative measures whose results are located only in behavioral domains rather than in complete descriptions of reading and writing. In a recent article in CCC O'Dell suggests, for example,



- 8 -

testing students writing for more than one purpose in order to determine whether the quality of writing is purpose-specific. He suggests that "whereas we once could use a single, widely agreed-upon procedure for evaluating all the writing done in a given mode, we may now have to use a variety of evaluative procedures, [he favors Lloyd-Jones' primary trait guides], most of which we have to create for ourselves." He continues by suggesting also the application of diverse tools for evaluation, but he nowhere suggests that the results generated by these tools might depend on a complex set of changing attitudes on the part of readers and students. While a writer may, for instance, more proficiently write a casual letter home than an essay to the general reader ~~persuading~~ against violence, a reader may not be able to judge the effectiveness of the letter at all if that reader was not its real audience.

I am calling additionally for methods much more congruent with the processes of writing and of reading. A situational method would correlate results from current measurements with the energy and intentionality of the acts of writing and reading. The motivation and the attitudes of both contributors to the meaning of a text would not be ignored. As a student, I knew my teachers valued improvement. I started slow. I knew that one teacher loved Strunk and White; I eliminated "the fact that." As a reader, I am prejudiced against the uses of "structure" as a verb and of small circles for the dots over i's. I expect Russell Baker to be funny, and I am willing to believe Bronowski's arguments supported by fewer examples than I demand of students writing to make the same points. In other words, I have written and read with varying degrees of authority and varying senses of not only my audiences' knowledge or prejudice about my subject, but also my audiences' knowledge and prejudices about me. I have evaluated my own work by how well it does all that I want it to do for my audience in a particular situation.

Since our new evaluative tools do not take into account these variables of writing and reading situations, much less the particular performance rules of simulations, which is what school and test writings always are, their results are dangerous in the hands of those whose minds may be empty of a sense of situation, of human interaction. We do not now have a complete calculus for evaluation, and we must acknowledge first that we do not. Although the inventory of tools we have grows, and seems to be growing in promising ways, we are not home yet. We would need to devise new descriptions of the affective as well as cognitive aspects of writing and reading, account for the predispositions of writer and reader, and set such measurements in varying ratios to the results from tools we now have before we were. If writing and teaching writing are to remain human activities but also become a coherent humanistic discipline, we need an interpretive theory informed by measurements of the variables in dynamic, organic situations. We might, for instance, investigate writers' evaluations of their own writing. Their opinions of its quality relative to their perceived abilities may, especially in extremely poor and extremely proficient writers, be an accurate control on other judgments suggested by text-measuring tools. A receptivity measurement for readers that identified their motivations, their prior expectations about the writer, and their expectations about the features of an appropriate text's response to a writing task might also be used to normalize other results.

Whether or not we have such additional measurements, we must at least begin to distinguish, clearly and insistently, between descriptive measurements of texts and qualitative judgments of the success of writing. Making this distinction is the only way to prevent traditional attacks on the brief impact of the English Department's teaching. And it will be the only way to forestall new criticism that is bound to result from teaching toward changes in, for example,

- 10 -

words per t-unit that do not really create change in a writer's ability to solve a complicated writing problem from the inventive subtext to the edited "publication."

The bottom line for teaching writing, one I do not find acknowledged in most school writing and testing programs, is, finally, that a good writer controls a variety of discourse situations, and has been taught not only a series of rules, but also the stated and hidden agendas of any particular writing situation. Rhetoricians would teach not only what a reader wants and needs from a text, but also the situational variability, even within academic writing situations, of those demands. Useful justifications for what we mean when we say a student wrote either well or poorly are ahead of us, and we should not rest now, with a new grammar, but without a complete pragmatics, of interpretation.